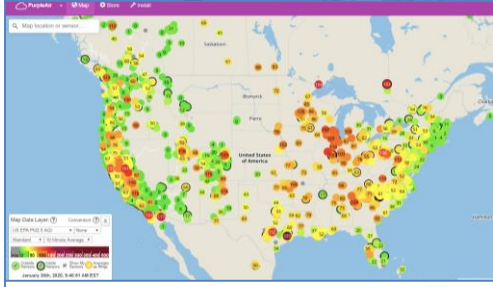
Air sensor data – What are the current technical practices and unmet needs of the EPA, state, local, and tribal air monitoring agencies?

Gayle Hagler, Ph.D.
Andrea Clements, Ph.D.
EPA Office of Research and Development

Why is the conversation needed now?

- Growth in the use of air sensor data to meet objectives at multiple levels of the government.
- Air sensor technology is evolving in a fragmented fashion, with variation in hardware, how data are handled, and business models.
  - Hardware: Pollutant types, other sensors (motion, meteorology), designed for stationary or mobile use
  - Data: Data storage on-board (or not), data transmission to server (or not), raw data or adjusted data (or both)
  - Legal / business: Data ownership by manufacturer, proprietary algorithms / intellectual property
- Understanding end user unmet needs on data management, analysis, and visualization is critical to design or advocate for optimal solutions.



Goals and scope

- Goal: Gain an understanding of the current practices, future outlook, and priority needs related to air sensor data management, processing, and visualization.
- Scope:
  - Target end user perspectives: EPA Regions, state, local, and tribal organizations
  - Level of sensor use ranging from minimal to fully integrated into meeting their organization's mission.
  - Timeframe for perspectives: Communicating present day situation, future outlook up to 5 years ahead

Approach

- Potential participants were recommended by EPA Region contacts and invited by EPA ORD to participate.
- Small group phone calls ranging 30 min to 1 hour were conducted as informal, open-ended dialogues. EPA ORD led the calls; several other EPA staff attended in listening mode to select conversations. Calls were not recorded; notetaking was conducted.
- Each phone call generally covered these topics:
  - Understanding organization's current level of use of air sensor data and technical practices
    - Included direct use of air sensors and any review/analysis of data from sensors operated by external parties
    - Included discussion of how they gained technical knowledge (sensors, data analysis approaches)
  - Understanding how the organization saw air sensors integrated into their next 5 years – discussed degree of growth and transitions in technical practices.
  - Discussing any unmet needs for effective use of sensor data by their organization, focusing on air sensor data but also capturing in notes any other related needs mentioned.

Approach

- Synthesis: Notes were tagged with the following labels; tagged text compared across dialogues and interpreted.

Current practices

- Level 1 User: limited use (e.g., educational demonstrations)
- Level 2 User: pilot-testing and evaluation of sensors
- Level 3 User: expansive use to meet organization's objectives
- Sensor Type
- Data Use Purpose
- Restricted Data Management
- Unrestricted Data Management
- Showing data to the public
- Data remains private
- Traditional Data Processing and Analysis: e.g., Excel
- Advanced Data Processing and Analysis: e.g., Script-writing (R, Python, etc.)

Future Outlook

- Low Growth Expected
- Moderate Growth Expected
- High Growth Expected

Unmet Needs

- Workforce Needs (i.e., training needs, more staff or staff time needed)
- Sharing of Technical Practices
- Sensor Products - Information and Performance
- New Data Management Solutions Needed
- New Data Analysis Tools or Functionality Needed
- Legal, Data Security or Data Ownership Issues
- Data Standards

Dialogue participants

- Talked with 19 organizations in total: 6 EPA Regions, 13 state/local/tribal air monitoring organizations
- EPA Regions (**R**): 1, 2, 4, 5, 7, 9
  - Regions were asked to share about their direct use of sensors as well as knowledge of use within their Region by air monitoring organizations.
- State/Local/Tribal organizations (**SLT**):
  - Kansas Department of Health & Environment
  - Iowa Department of Natural Resources
  - California Air Resources Board
  - South Coast Air Quality Management District
  - Imperial County Air Pollution Control District
  - Sacramento Metropolitan Air Quality Management District
  - Georgia Environmental Protection Division
  - South Carolina Department of Health and Environmental Control
  - North Carolina Department of Environmental Quality
  - Mecklenburg County Air Quality
  - New York State Department of Environmental Conservation
  - New Jersey Department of Environmental Protection
  - Confederated Tribes of the Colville Reservation

Summary: Level of Use and Future Outlook

- **Organizational type was a poor predictor of level of sensor use and unmet technical needs** – significant variation among Regions and SLTs.
  - Organizing into Level I, II, and III level of users discovered commonalities in purposes of sensor data use and technical needs
  - Capacity is another key factor to organize needs
    - Staff time available
    - Staff skill sets for data analysis and programming

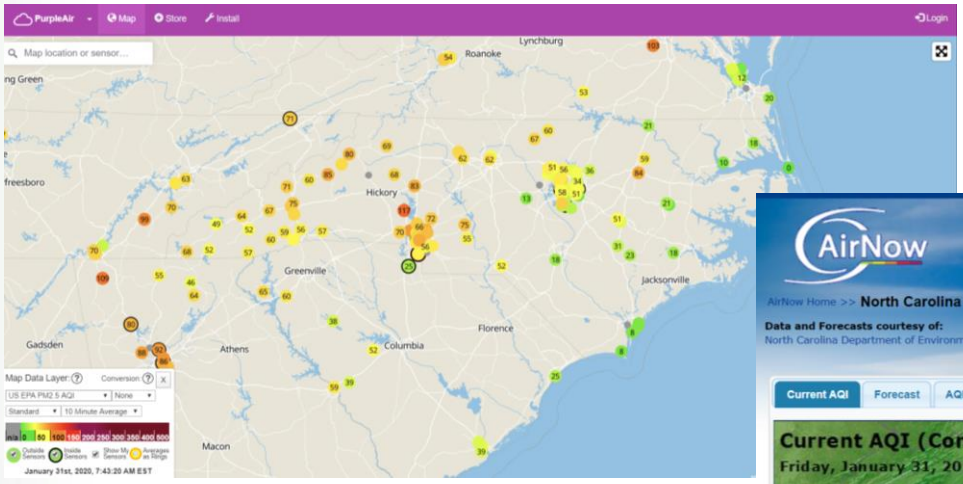
- No one expects low growth – **all are anticipating sensors to increasingly play a role** in their organization's objectives
  - Responding to, informing, and loaning equipment to the public
  - Use of sensors for screening, temporary monitoring, investigative purposes, and continuous monitoring

- Most organizations **do not expect additional resources** (staff or funds) to support the additional workload related to sensors.

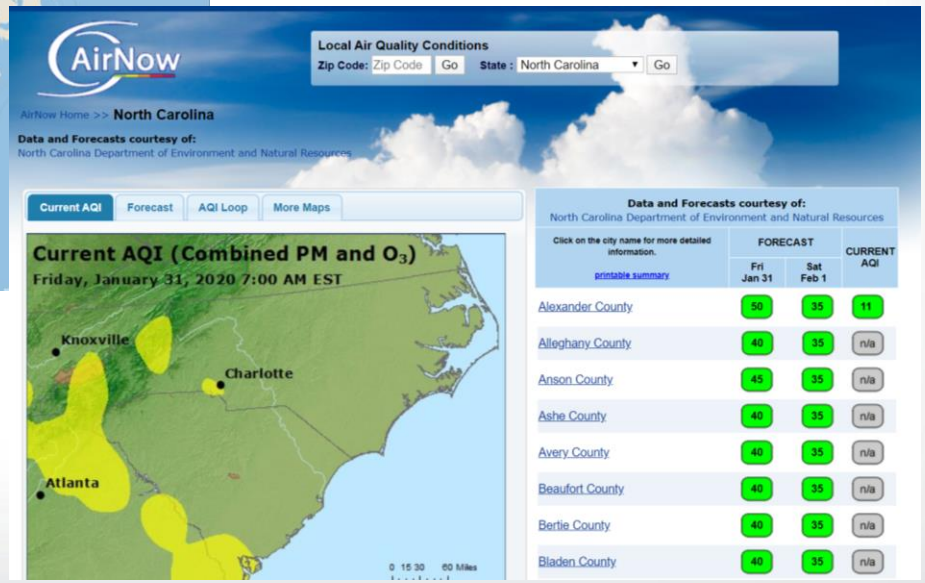
Summary: Level of Use and Future Outlook

- **Degree of use varied widely**
  - We categorized the 19 organizations as: **Level I (5)**, **Level II (11)**, **Level III (3)**

Level I example: Keeping an eye on the space and occasionally viewing public-facing sensor data on a website



e.g., quick scan to compare of public-facing information



- Cursory level look for better understanding
- Data are not directly accessed nor analyzed in-depth

Summary: Level of Use and Future Outlook

Level II example: Use of sensors for temporary monitoring to assess a citizen complaint, investigate an area of concern, and/or test quality of sensor data by collocating with reference monitor.

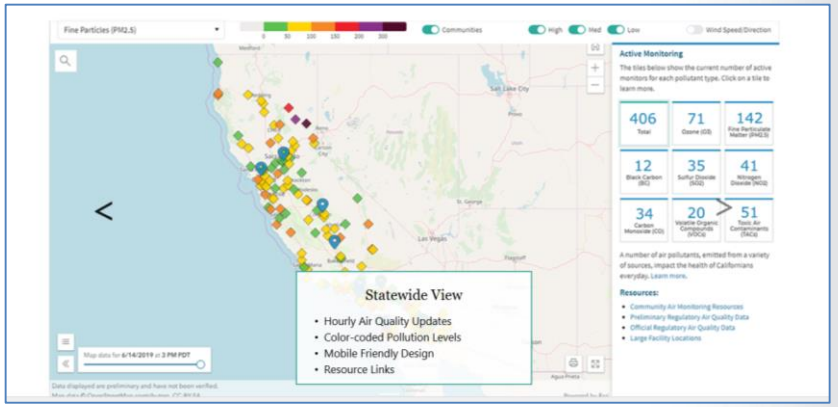
- Typically limited number (<10) in use; temporarily collecting monitoring data.
- Data are accessed and analyzed
- Data may be made public or kept private depending on application.
- Many rely on on-board memory and manual data download; fewer use online sensors (data to manufacturer or private server)

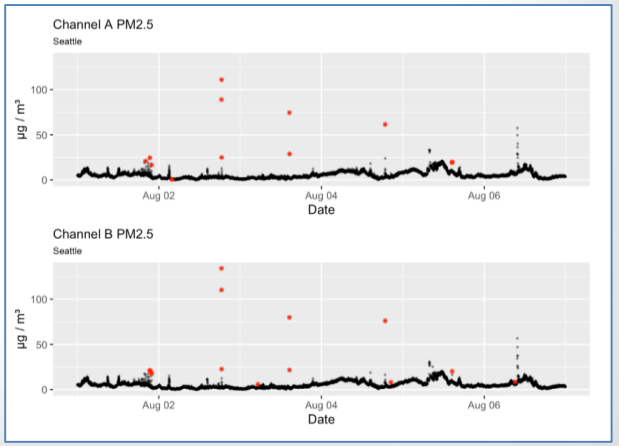
Summary: Level of Use and Future Outlook

Level III example: Sensors integrated into meeting organization's goals, used in greater number and for longer periods of time. Currently only several California organizations were at this level, all cited AB617 as one important driver.

- High number of sensor data sets of interest to analyze.

- Large data volume demanding new methods for data management, data visualization, and data analytics.
  - Two notable systems under development or recently created:
    - AQView – data storage and visualization platform supporting AB617, allows multiple pollutant and instrument/sensor types
    - AirSensor R package – currently focuses only on PurpleAir sensors





Summary: Current Practices

Sensor data-related issues Level II and Level III users are currently grappling with throughout implementation

- Sensor Selection
  - Online sensors (cell, wifi)
  - Offline sensors (on-board data storage)

Cost of equipment and data management

Can this equipment interface easily with my existing data management system?

Can I handle the workload of manual data downloads?

Information security limitations

Am I allowed to use wifi at the deployment location?

Am I allowed to have my organization's data stored on the manufacturer's cloud / server?

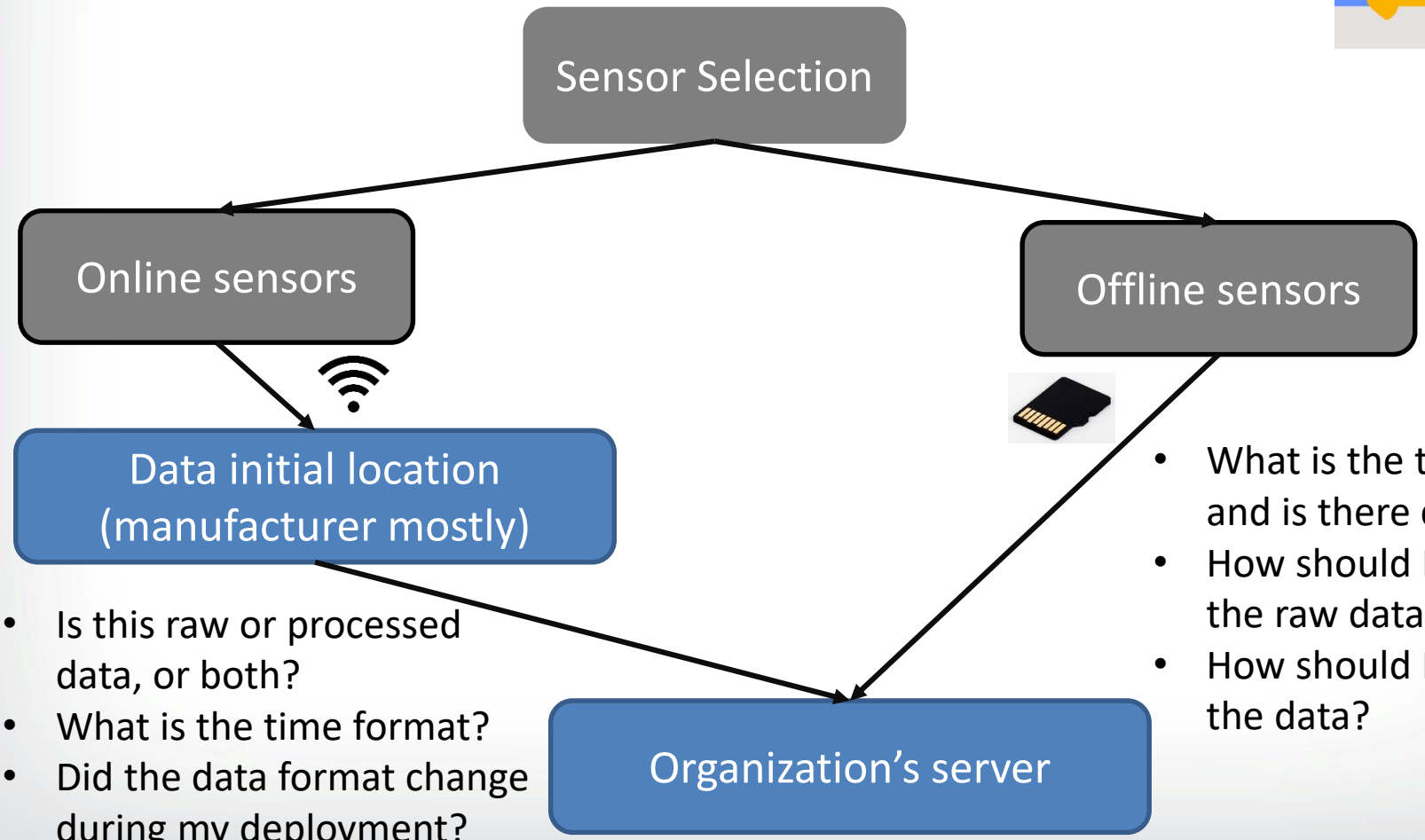
Limits imposed by organization

Do I trust how a manufacturer would treat the data? (ownership, algorithms, data standards, data display)

Limit imposed by user

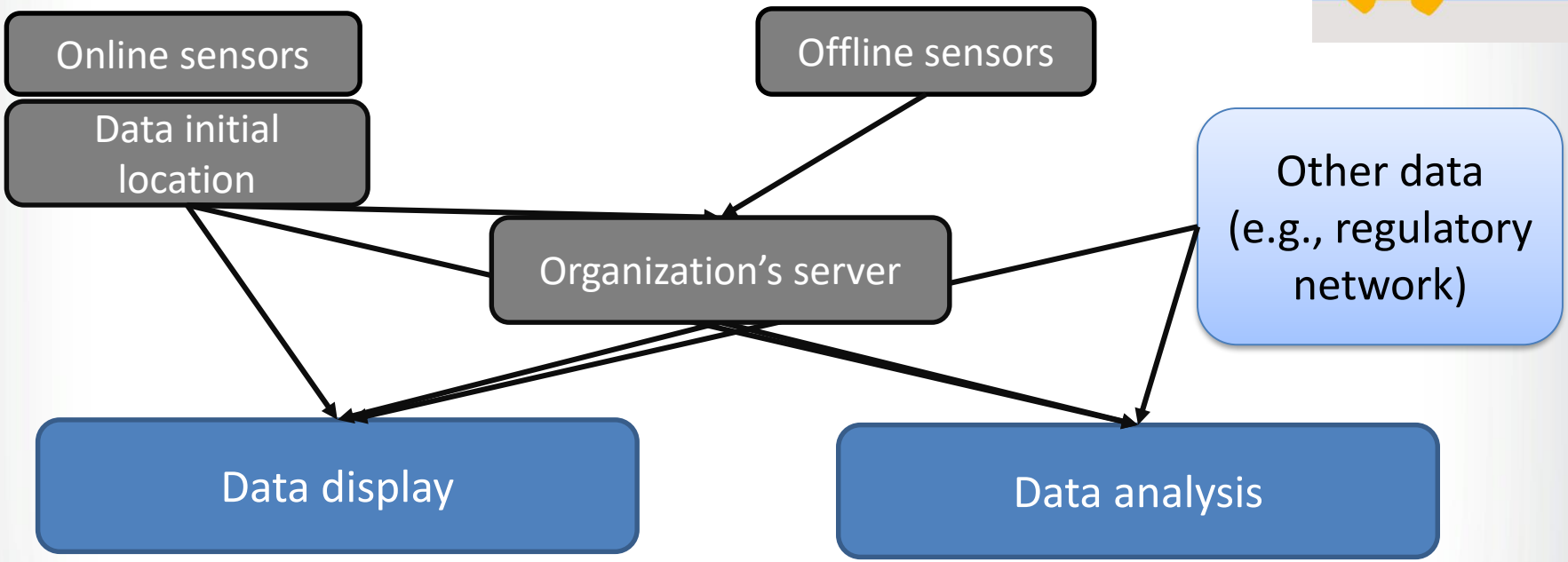
Summary: Current Practices

Sensor data-related issues Level II and Level III users are currently grappling with throughout implementation



Summary: Current Practices

Sensor data-related issues Level II and Level III users are currently grappling with throughout implementation



- Do I like the display / analysis tools provided by the manufacturer? (if using manufacturer cloud or server)
- Are there staff members with time / skills to do advanced data analysis?
- Do we have resources to develop custom solutions?

Summary: Current Practices

User Level	Data Storage	Data Analysis / Processing	Current pain points
I	None – sensor use limited	No in-depth analysis – just quick viewing on a website or instrument integrated screen.	• No easy way to visually compare sensors with nearby reference monitors • Access to recent information on sensors, projects, best practices
II	Mixed: 1) On-board storage and manual download OR 2) Dependent on manufacturer server	Mixed: 1) Excel or other spreadsheet-based analysis (e.g., JMP) OR 1) Custom scripts developed by user (R most common followed by Python)	• Lack of staff time available • Time burden of data management and analysis • Learning curve and time burden to do custom-scripting (“everyone is an amateur coder”) • Lack of data standards • Insufficient information from sensor manufacturers on data
III	Custom-built data management system; cloud-based or transitioning to cloud; some use manufacturer servers	Custom-developed scripts and user interfaces (R, Python)	• Cost of cloud-based data management • Data provenance / data standards • Data security • Making data understandable to public

Summary: Unmet Needs (1 of 5)

New data management solutions (Level II and Level III users)

- Quote from an EPA Region participant: A manufacturer server “"might be fine if it was my personal sensor data going to a manufacturer cloud/website, but if it is EPA data that wouldn't fly."
- Many desired a dedicated location to store data
- AQS and AirNow/AirNow-Tech both mentioned as examples – desire system that supports real-time data storage, some level of public display, analysis capability.
- Want government-hosted data storage – concerns about data security and control of data, legal issues, possibility of data manipulation
- A few mentioned they would be comfortable with non-government server if sufficiently documented, chain-of-custody for data, and guidance provided by EPA.

Summary: Unmet Needs (2 of 5)

Data standards are needed (voiced by Level II and Level III users)

- No standard format
- No standard tracking of data provenance / history
- Lost time developing custom scripts when sensor firmware updates result in data format change.
- Lack of data standards is particularly problematic for organizations pooling sensor data from multiple entities.

Summary: Unmet Needs (3 of 5)

Automated and quick viewing of sensor data integrated with other data (regulatory monitors, other sensors, meteorology) – Level I, Level II, and Level III Users

- Low technical barrier to use (non-programmer friendly)
- Allows offline data to be manually added
- Supports some level of data screening
  - Outlier detection / flagging
  - Application of data correction schemes if available for a particular sensor
- Supports geospatial exploratory analysis
- Supports combined air pollution and meteorology analysis
  - Wind flow patterns
  - Pollution-wind comparisons (e.g., pollution roses)
  - Measurement artifacts (e.g., PM sensor data vs. relative humidity)
- Avoids implication of endorsement (of the data, of the sensor types)
- Training support provided (videos)

Summary: Unmet Needs (4 of 5)

Code-sharing and code development support (Level I, Level II, and Level III Users)

- "Everyone is working on writing code in their own little area".
- R cited most frequently as the code being used, followed by Python.
- As sensor use increases and organizations transition from Level 1 to Level 2 use, migrating toward custom coding to handle data.
- Only Level 3 organizations have data scientists on staff supporting use of sensors.
- One EPA Region mentioned potentially standardizing use of R (currently in higher level use in water).

Summary: Unmet Needs (5 of 5)

More sharing of technical practices desired (voiced by Level I and Level II Users)

- Forum or portal for communication among SLTs – share real-world experiences, Q&A
- EPA Regions also interested in communication with other Region – similar sharing of real-world experiences, Q&A about sensors.
- Sharing of best practices and knowledge – share how sensor data are managed and quality assured.
- Training sessions on sensor data management / analysis

Next steps

- These findings will be presented internally at EPA to support planning discussions.

- These findings will also be shared in a public conference, hopefully stimulating further conversation and sharing of ideas to address unmet needs.

EPA ORD point of contact:
Andrea Clements
clements.andrea@epa.gov

Thank you!!

Dialogue Participants from these organizations:

- EPA Regions: 1, 2, 4, 5, 7, 9
- State/Local/Tribal organizations:
  - Kansas Department of Health & Environment
  - Iowa Department of Natural Resources
  - California Air Resources Board
  - South Coast Air Quality Management District
  - Imperial County Air Pollution Control District
  - Sacramento Metropolitan Air Quality Management District
  - Georgia Environmental Protection Division
  - South Carolina Department of Health and Environmental Control
  - North Carolina Department of Environmental Quality
  - Mecklenburg County Air Quality
  - New York State Department of Environmental Conservation
  - New Jersey Department of Environmental Protection
  - Confederated Tribes of the Colville Reservation

Appendix

More Details from Dialogues Organized by Level of Use and Topic

Summary by User Level: Level 1

- 5 participants were tagged as Level 1 Users (low current use of sensors and sensor data)
- **Current Practices**
  - R: Currently monitoring what is happening with low-cost air quality sensors and seeing what state and local are doing with them.
  - SLT: Currently not using low-cost sensors of their own, but partnering with communities and companies to evaluate, correlate, correct, and collocate sensors.
  - R: Doing visual checks on various websites but no in-depth analysis."
  - SLT: Developing guidance for the public on how to use low cost air sensors which will eventually be posted on their website.
  - SLT: IT doesn't allow their device to go on their wifi – using memory cards and manual download of data.
- **Future Outlook: Low to Moderate Growth**
  - Some felt their use level would be the same in 5 years, others felt it would be growing (e.g., initiating a sensor loan program, replying to more questions by the public)

Summary by User Level: Level I

- **Unmet Needs:**
  - **Automated and quick viewing of sensor data integrated with other data (regulatory monitors, other sensors, meteorology data)**
    - R: User friendly - upload the data and not require data manipulation by the user
    - R: Shows graphs like time series, maps if there were GPS coordinates
    - R: Sensor vs. regulatory data comparison, including a visual AQI color comparison
    - SLT: Tools to visualize, correct, and manage the data for sensors with large market share or a single portal where everyone shares data in a standardized way.
    - R: Combine in wind data visualization - interest in ground-level winds; but also exceptional events looking at larger scale transport
    - R: Show current conditions; recent history (back a year or two) adapting to any user-provided data
    - R: Would include manufacturer's web services data (API pull) and also allow user to manually upload data
    - R: Data presentation needs to be caveated - need to avoid implication of endorsement.
  - **Data management**
    - R: AirNow / AirNow-Tech as a model

Summary by User Level: Level I

- **Unmet Needs:**
  - **Approach for code-sharing (R, Python) for programmers**
    - R: "Everyone is working on writing code in their own little area".
    - R: Code-sharing would make life easier. But that requires more expertise.
  - **Information sharing**
    - SLT: Share information to see where current research is, highlights of projects happening across the state, and pitfalls to avoid
    - SLT: Share how different projects manage their data and quality assurance
    - R: Would like a better way to coordinate between each other for large-scale events - that goes across boundaries. They have regular calls but would like to see a portal for sharing information.
    - SLT: Online resources are most useful, as are webinars and research presentations
  - **Skill-building**
    - R: Can someone help with code troubleshooting?
    - R: "Everyone is an amateur coder and gets stuck, then is stuck for a long time"
  - **Other:**
    - Work on educational materials needed

Summary by User Level: Level II

- 10 participants were tagged as Level II Users (technology testing and moderate use of sensor data)
- **Current Practices**
  - R: Plan to use sensors to track woodsmoke
  - R: 17 air sensors located for a research study in a city
  - R: Cell modems are a significant financial burden; manually collecting the data
  - SLT : Have 3 PurpleAir sensors collocated with FEM monitors
  - SLT : Has 4 sensors that are used for public complaints to collect supporting data
  - SLT: Own 25 PurpleAir sensors, currently using 9. Inside and outside monitoring.
  - SLT: Have been pulling data files from [sensor] website, and then using R for descriptive statistics and excel to regress data against regulatory monitors and correct for relative humidity and temperature.
  - R: Considering standardizing in R since much of the Region uses R
  - SLT: Analysis done with excel, 24 hour averages
  - SLT: Doing data analysis in excel using macros for rounding and offsets

Summary by User Level: Level II

- **Current Practices (continued)**
  - SLT : Built custom scripts to pull data from PurpleAir - 20+ sensors
  - SLT : Data is publicly available on sensor manufacturer's website and used without correction for public information. Emphasize trends over absolute values.
  - SLTs: Differing options on [sensor manufacturer] hosting/displaying data – some frustrated with it, some satisfied by it.

- **Future Outlook: Moderate to High Growth**
  - R: Expand use for woodsmoke monitoring, informing siting of regulatory monitors
  - R: Start an air sensor equipment loan program
  - R: Considering using as a screening tool of air quality in different areas.
  - SLT: Schools are already buying sensors and they are expecting future growth
  - SLT: Tribes are buying 1 or 2 sensors and then adding as their skill expands
  - SLT: Future use in investigative purposes.
  - SLT: Developing project for drones with sensors.
  - R: Do not plan on acquiring lots of sensor hardware themselves, but will work with data and act as a technical advisor to other organizations/regions
  - SLT: More sensors to come as AB 617 and air district are adding them separately

Summary by User Level: Level II

- **Unmet Needs**
  - **Automated and quick viewing of sensor data integrated with other data (regulatory monitors, other sensors, meteorology data)**
    - R: Would like better tools to analyze their own data as well as publicly available data
    - R: Data system that would provide real-time data storage, separate but similar data system as AQS, allows comparison of regional citizen science data with reference air quality data, publicly viewable.
    - R: We have lots of data, but few staff have the coding experience handle it.
    - R: Need better tools for sensor performance checking that can be used for all sensor types and that can be used by those without much coding experience
    - R: Would like to be able to include data from other organizations into their studies so long as there was tracking as to where the data came from.
    - SLT: Provide historic trend and current conditions
    - SLT: Include weather station data including wind speed and direction.
    - SLT: Need tools for screening the data.
    - SLT: Would like better forms of geospatial, statistical and graphical analysis.
    - STL: Might like to look at higher time resolution data.

Summary by User Level: Level II

- **Unmet Needs**
  - **Data management**
    - SLT: Want to compile all the sensor data but keep it separate from regulatory data
    - SLT: Looking for better method of data management, eventually to the same format as regulatory type data
    - SLT: Using [data management system] to store regulatory data, but it is too expensive to also store the sensor data
    - R: Want quality check information and other metadata kept with the data
    - R: Cautious about public sharing of low-cost sensor data, about potential corporate manipulation of data, and of data leaving EPA control
    - R: See legal issues with using manufacturers dashboard for data storage
    - R: Can't use vendors system for storage due to requirements, but comfortable with private servers if sufficiently documented, chain of custody assuming formal agreement with manufacturers or guidance from OAQPS
  - **Approach for code-sharing (R, Python) for programmers**
    - R: Current bottleneck in data processing and visualization Only doing basic analysis due to constraints Concerned about scale-up effort.

Summary by User Level: Level II

- **Unmet Needs**
  - **Information sharing**
    - SLT: Share information to see where current research is, highlights of projects happening across the state, and pitfalls to avoid
    - SLT: Share how different projects manage their data and quality assurance
    - R: Would like a better way to coordinate between each other for large-scale events - that goes across boundaries. They have regular calls but would like to see a portal for sharing information.
    - SLT: Online resources are most useful, as are webinars and research presentations
  - **Skill-building**
    - SLT: Using interns to do coding and analyze the data.
    - SLT: Don't have people to do more work, but sensor technology is attractive because it is perceived to be less expensive and lower maintenance.
    - SLT: If someone could add resources to handle the data, this would reduce strain.
    - SLT: Need data analysis to take up less time in the future.
    - R: We need someone focused on air sensors and citizen science.
    - R: We have few staff with coding experience.

Summary by User Level: Level II

- **Unmet Needs**
  - **Data standards**
    - R: Hoping for transparent information about algorithms and chain of custory of data.
    - SLT: Issues with getting data conversions from other organizations and cannot replicate.
    - SLT: Issues with timestamps for sensor data.
    - SLT: Want more information on sensors themselves – often run into time stamp issues and having to do a lot of custom coding.
    - R: Data isn't consistent between vendors.
  - **Other:**
    - S: Want more learning materials from sensor manufacturers.
    - R: Need educational materials and guidance for the public.
    - S: Want more educational information about sensor maintenance and how local sources can affect the measurements.

Summary by User Level: Level III

- 3 participants were tagged as Level III Users (sensors fully integrated into meeting organization's goals)
- **Current Practices**
  - SLT: Funding for AB617 communities to deploy their own sensors
  - SLT: Sending sensors to air districts with wildfires
  - SLT: Goal of increasing public awareness of air quality, determining pollution distribution, and determining sources.
  - SLT: Drivers for increase in sensor use include AB617, Rule 1180, and EPA grants
  - SLT: Using sensors for indoor and outdoor sampling during wildfire smoke events
  - SLTs: Building custom data management solutions
  - SLT: Using manufacturer servers for some sensor types
  - SLTs: Building custom public-facing display of sensor data
  - SLTs: R package developed for sensor data analysis
  - SLT: Pulling data together in many ways, sometimes include data scraping from websites.
  - SLTs: Using R and Python for data analysis and visualization
  - SLT: Have access to data scientists and programmers
  - SLT: Use RETIGO for some visualization

Summary by User Level: Level III

- **Unmet needs**
  - **Automated and quick viewing of sensor data integrated with other data (regulatory monitors, other sensors, meteorology data)**
    - SLT: Pulling data together in many ways, sometimes that includes data scraping from websites
    - SLT: Would like more tools for coupling meteorological data with sensor data
    - SLT: First goal is to achieve general statistics, and then will eventually need geospatial visualization
    - SLT: Better visualization, communities like calendar plots (openair plot option); there are issues with AQ map colors, and wind roses are hard to understand
    - SLT: Want to be able to tell community why sensor results are the way they are (whether error or actual phenomena)
  - **Data management**
    - SLT: They are often using adhoc DIY solutions
    - SLT: A way to convert data from all sensors/manufacturers into the same data type
    - SLT: Looking for better means of data management (perhaps cloud-based storage) and data security
    - SLT: Cloud-based data management is quite expensive, looking to cut costs
    - SLT: Many communities want control of their own data

Summary by User Level: Level III

- **Unmet needs**
  - **Information sharing**
    - SLT: Want performance standards by which to judge sensor data
  - **Data standards**
    - SLT: Don't know the accuracy of data since some districts are adjusting the data before posting, and some sensors are mislabeled. Would like to add metadata to include provenance, data lineage, etc.
    - SLT: Want agreement on meta-data; some criteria to define provenance/heritage, QA/QC, version of firmware
  - **Other:**
    - SLT: Want to better communicate information to the public, share a similar message from district to district
    - SLT: Want a "sensor toolbox" for local agencies
    - SLT: Data security is an issue as one of their air monitoring stations was recently hacked
    - SLT: Looking for a way to audit devices once they are deployed in the field.